A Queueing Model of a Spread Spectrum Multiple Access*

Vladimir Marbukh and Nader Moayeri

National Institute of Standards and Technology
100 Bureau Drive, Stop 8920, Gaithersburg, MD 20899-8920 USA
e-mail: marbukh@nist.gov, moayeri@nist.gov

Abstract

We propose a queueing model of the reverse link from mobiles to the base station in a single cell of a cellular DS-CDMA system. Approximation of independent queues leads to a non-linear fixed point equation that describes globally or locally stable system modes. We argue that interactions between queues at different mobiles through interference at the base station are essential for the system queueing performance including traffic delays and backlogs. To account for these interactions we propose to model DS-CDMA by the enhanced Generalized Processor Sharing (GPS) scheduling discipline. It appears that this model strikes a compromise between accuracy and tractability. At least some already known performance results for GPS can be directly used or adjusted to provide performance estimates for DS-CDMA.

1. Introduction

The main force driving evolution of wired networking technology is the need to handle multimedia traffic with widely varying statistical characteristics and with specific requirements for quality of service (QoS). While this trend will continue, future network infrastructures will be a mixture of both wired and wireless networks. In fact, wireless access to the Internet will probably become much more common than wired access. To prevent waste of resources in the wired backbone, the wireless access network must have the same or comparable capabilities to handle multimedia traffic. Today, wireless technology lags far behind wired technology in its ability to serve multimedia traffic. The major reason for that is restrictions on the wireless technology imposed by wireless channel impairments and by severe limitations on wireless bandwidth and on transmission power by mobiles.

Since ARQ is limited in its ability to stream multimedia traffic, issues such as bandwidth sharing, power efficiency, and transmission robustness must be addressed by the design of a medium access control (MAC) protocol. The performance of MAC protocols largely determines the ability of a wireless network to handle real-time multimedia traffic. Future MAC protocols will likely employ Spread Spectrum Multiple Access (SS-MA). The inherent traits of SS-MA to combat interference and to provide a degree of protection against jamming and interception, as well as power efficiency, has caused the military to make wide use of SS-MA for message traffic. Now these same traits, in addition to soft capacity limits and to an ability to provide bandwidth on demand, are causing SS-MA to gain popularity for commercial use. The future commercial success of SS-MA will depend largely on its ability to handle multimedia traffic.

Up to now, commercial wireless technology, used mainly for exchanging voice calls, has relied primarily on time-division multiple access (TDMA) to share wireless channels. Recently proposed TDMA enhancements, such as Packet Reservation Multiple Access (PRMA), introduce some statistical multiplexing capabilities into TDMA, making competition between TDMA and SS-MA much closer. Enhanced TDMA systems are capable of directly allocating time slots on demand at the cost associated with making reservations. SS-MA offers bandwidth on demand without need for a reservation. However, in SS-MA a user's bit-service rate is a complex non-linear function of various intrinsic parameters, such as transmission power and processing gain, as well as various extrinsic parameters, such as interference from other users and conditions on the wireless channel. To provide guaranteed QoS for multimedia traffic, SS-MA requires fast, closed-loop control of system parameters based on real-time information about the intrinsic and extrinsic parameters, and about the performance of the system. Carrying out the following program would provide theoretical foundation for

* This work was supported in part by a grant from the NIST Advanced Technology Program (ATP)

developing SS-MA based technology able to carry multimedia traffic:

(1) development of analytical and simulation performance models of SS-MA,
(2) identification of system parameters, besides transmission power, having the greatest effects on the provision and maintenance of QoS for multimedia traffic,
(3) development of control algorithms based on the parameters identified, and
(4) comparison of various wireless systems against each other and against theoretical limits.

Despite some partial results along the lines of this program, much work remains to be done. This paper is addresses step (1), which will provide a basis for understanding SS-MA and the subsequent steps. We concentrate on a Direct Sequence Code Division Multiple Access (DS-CDMA) system. DS-CDMA is the technology for the 3rd and probably following generations of mobile systems. The paper is organized as follows. Section 2 briefly describes a packet based DS-CDMA. Section 3 discusses approximation of independent queues at different mobiles. This approximation leads to a fixed point equation that describes globally or locally stable system modes. In section 4 we argue that interactions between queues at different mobiles are essential for the system queueing performance including traffic delays and backlogs. To account for these interactions we propose to model DS-CDMA by the enhanced Generalized Processor Sharing (GPS) scheduling discipline. It appears that this model strikes a compromise between accuracy and tractability. At least some of the already known performance results for GPS can be directly used or adjusted to provide performance estimates for DS-CDMA. Finally, section 5 gives some examples of these results.

2. A Packet Based DS-CDMA

Fig. 1 shows a traffic flow in the reverse link from user j to the base station in a packet-based DS-CDMA.

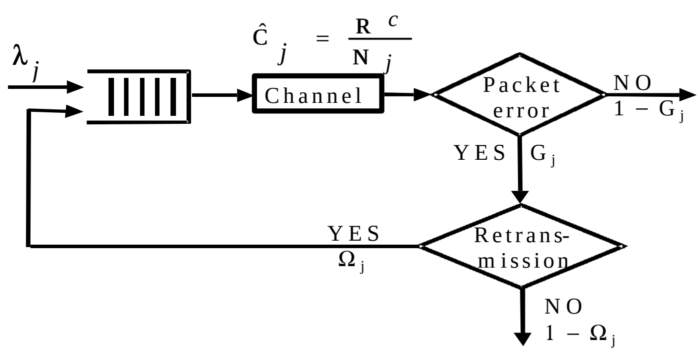

Figure 1: Traffic flow in DS-CDMA

User j generates traffic at an average rate $l_j[bits/\sec]$. This traffic enters a buffer of size $B_j[bits]$, if buffer space is available. Traffic from the buffer is transmitted at the constant rate

(1) $\hat{c}_j = R^c \big/ N_j$,

where R^c is the chip rate of the system and N_j is the user j processing gain. Transmission rate (1) is also the peak bit service rate for user j. Due to interference, only a fraction of transmitted packets is accepted at the base station as received correctly (i.e. without errors, or with an acceptable number of bit errors). These packets leave the system as successfully delivered. Let $1-G_j$ be the probability that a transmitted packet is successfully delivered. Packet error probability G_j depends on the bit energy-to-noise ratio, coding scheme, etc. Some packets that are received with an unacceptable number of bit errors are retransmitted. We model this feature of the system assuming that each packet received with unacceptable number of bit errors is retransmitted with probability Ω_j or leaves the system with probability $1-\Omega_j$. We assume, for the purpose of estimating performance, that G_j and Ω_j remain constant regardless of how many times a packet is transmitted until it is either received with an acceptable number of errors or it leaves the system. We also assume that a packet stays in the buffer until it leaves the system. For brevity we consider a case of infinitely large buffers: $B_j = \infty$.

It is easy to see that given G_j, the average clearing rate for traffic backlogged at mobile j is:

(2) $$c_j = \hat{c}_j(1 - \Omega_j G_j).$$

Since retransmissions introduce additional packet delays, for real-time traffic with very stringent delay requirements no retransmission is allowed ($\Omega_j = 0$), and the clearing rate is constant and equal to the peak service rate: $c_j = \hat{c}_j$. For non real-time traffic, with very loose delay requirements, there is no restrictions on the number of retransmissions ($\Omega_j = 1$) and the average clearing rate is: $c_j = \hat{c}_j(1 - G_j)$. The packet error probability G_j is a decreasing function of the user j bit energy-to-interference ratio $(E_b / I_0)_j = N_j SIR_j$:

(3) $$G_j = G_j[(E_b / I_0)_j],$$

where N_j is the processing gain and SIR_j is the Signal-to-Interference Ratio for user j. We make the common assumption [1] that

(4) $$SIR_j = \frac{P_j x_j}{s^2 + \sum_{i \neq j} P_i x_i d_i},$$

where P_i is the user i transmission power, x_i is the power attenuation factor from user i to the base station, $d_i = 1$ if user i transmits and $d_i = 0$ otherwise, and s^2 is the average power of the background noise at the base station.

3. Approximation of Independent Queues

Combining (1)-(4) we can express user j bit service rate (2) as a function of the signal-to-interference ratio SIR_j: $c_j = c_j(SIR_j)$. The problem, of course, is that service rate $c_j = c_j(SIR_j)$ for user j depends on congestion at other mobiles $i \neq j$ through indicators d_i. As a result of these interactions, all queues at different mobiles are correlated and should be considered jointly. Unfortunately, this system of interacting queues is too complex to allow for performance evaluation without some additional assumptions and simplifications. Obviously, $\Pr\{d_i = 1\} = E[d_i] = r_i$ where the average utilization of the wireless channel from mobile j to the base station is:

(5) $$r_j = \min\{1, l_j \big/ c_j^{ave}\},$$

and the average bit service rate for mobile j is $c_j^{ave} = E[c_j(SIR_j)]$. The most radical simplification is to assume that each user j gets a fixed bit service rate. This simplification effectively decomposes the system into the collection of independent queues. Then, given fixed service rates $c_j^{ave} = E[c_j(SIR_j)]$, performance of each of these queues can be easily evaluated. Under this approximation, indicators d_i are jointly statistically independent and the average service rates is:

(6) $$c_j^{ave} = \sum_{d_i} c_j\left(\frac{N_j P_j x_j}{s^2 + \sum_{i \neq j} P_i x_i d_i}\right) \prod_{i \neq j} r_i^{d_i} (1 - r_i)^{1 - d_i}.$$

Substituting (5) into the right-hand side of (6) we obtain a system of J non-linear fixed point equations for the average utilization r_j, $j = 1,..,J$. If the number of mobiles is large, i.e., $J >> 1$, solving this system becomes difficult. Fortunately, if $J >> 1$ and the average power received at the base station from any one mobile is a small portion of the total average power received from all mobiles, then:

$$(7) \qquad r_i P_i x_i << P^+ = \sum_{j=1}^{J} r_j P_j x_j, \quad i = 1,..,J \,.$$

The system of J fixed point equations (5)-(6) can be approximately replaced with a single fixed point equation [2]. This drastic reduction in the dimension of the system description is attributed to the law of large numbers that suggests that under (7) the total power received at the base station from all mobiles is very close to its average value and consequently,

$$(8) \qquad P^+ = \sum_j x_j P_j \min\{1, \quad l_j \Big/ c_j \big(\frac{N_j P_j x_j}{s^2 + P^+}\big)\}.$$

After solving equation (8), the average utilization of the reverse link from mobile j to the base station can be easily estimated as follows:

$$(9) \qquad r_j \approx \tilde{r}_j = \min\{1, \quad l_j \Big/ c_j \big(\frac{N_j P_j x_j}{s^2 + P^+}\big)\}.$$

It can be expected that globally (locally) stable solutions of non-linear fixed point equation (8) describe globally (locally) stable systems modes. Quantitatively, possible presence of multiple locally stable system modes can be explained as follows. A temporary congestion at some group of users due to sudden burst of the incoming traffic or temporary wireless channel fading, negatively affects the throughputs and increases probability of congestion for other users through increase in interference at the base station. Then this newly created congestion can reinforce the initial congestion. Due to this positive feedback, temporary congestion, and consequently packet delays and buffer overflows tend to occur synchronously at the different mobiles. As a result of this mechanism the system as a whole can be driven to the congested mode.

If in some persistent system mode average utilization $r_j < 1$ ($r_j = 1$), then the user j experiences finite (infinite) traffic delays. The space of system parameters, i.e., transmission powers, processing gains, etc., can be partitioned into regions with topologically different phase portraits of equation (8) and different sets of congested users. This partition represents the phase diagram of the DS-CDMA system and can serve as a very useful tool for qualitative description of the system. Large fluctuations cause rare transitions of the system from one locally stable mode to another. Small fluctuations leave the system in the current, locally stable state and are responsible for the traffic delays and backlogs. Phase diagrams and transitions between locally stable modes have been discussed in [3] with respect to wired networks. These results can be adjusted to a case of DS-CDMA system.

4. Approximation of Interacting Queues

It can be expected that an approximation of independent queues adequately describes globally or locally stable system modes, and depending on whether the average utilization $r_j < 1$ or $r_j = 1$, predicts stability of the session j in some particular persistent system mode. Also, the average utilization $0 < r_j < 1$ gives some qualitative indication of the queueing performance for user j in the corresponding persistent mode. However, since an approximation of independent queues does not take into account that in DS-CDMA all users share the same bandwidth, this approximation cannot be used for quantitative conclusions about the queueing performance, including delays and backlogs. To stress this critical point, consider a case of two mobiles $J = 2$ with significantly different received transmission powers: $P_1 x_1 >> P_2 x_2$. In this case an adequate performance model is a priority queueing system where first queue has priority over second queue in terms of access to the channel. The performance of this priority queueing system is very much different from performance of the system with dedicated channels. This extreme case illustrates the following important point. In presence of multimedia traffic with significantly different Quality of Service (QoS) requirements and statistical characteristics, a user queueing performance is determined by interactions between queues due to interference at the base station.

There is an obvious trade-off between accuracy and tractability of the queueing model of a DS-CDMA. In order to be relevant to DS-CDMA a queueing model has to meet the following minimum set of requirements. The model has to assume statistical multiplexing on the channel (due sharing the same frequency band by different mobiles) and contain the following set of controlled parameters:

- Controlled peak service rate to account for user specific bit service rates (1), and
- Controlled sharing of the bandwidth among different users to account for controlled sharing of the same frequency band by different users. The user bit service rate (2) depends on the user characteristics, including transmission power, processing gain, coding scheme, etc.

We propose a following particular form of piece-wise linear approximation of the bit service rate for user j:

(10) $$c_j = \min\{\hat{c}_j, R_j SIR_j\},$$

where the peak rate $\hat{c}_j$ is given by (1) and parameter R_j depends on the user characteristics as well as the particular persistent system mode. In a case when system parameters, including processing gain, coding and modulation schemes, are extremely adaptable and all users have the same packet retransmission probabilities ($\Omega_j = \Omega \quad j = 1,..,J$) the bit service rate $c_j(SIR_j)$ for user j can be approximated as follows:

(11) $$c_j = R * SIR_j .$$

A model (12) is a particular case of (11) for sufficiently high peak rates $\hat{c}_j \to \infty$, and user independent parameters $R_j = R, j = 1,..,J$. For example, in an ideal DS-CDMA system the bit rate for user j is: $c_j = W \log_2(1 + SIR_j)$, where W is the spreading bandwidth. Consequently, for a model (11) to be physically feasible, parameter R should satisfy the following condition: $R \le R^{\max} = W/\ln 2$.

The major advantage of the approximation (10) is that it strikes a reasonable compromise between the accuracy and tractability of the queueing model. On the one hand, model (10) meets the minimum set of requirements. On the other hand, model (10) appears to be tractable as a natural enhancement of the Generalized Processor Sharing (GPS) scheduling discipline. The GPS was introduced in [4]-[5] and then extensively studied under various traffic conditions (see for example [6]). GPS is a work conserving scheduling strategy that can be regarded as the limiting form of weighted round robin policy with very short time slots and traffic is treated as an infinitely divisible fluid. Extensions to account for the finite packet size are not difficult and will not be considered here. The GPS is defined as follows. Let $J+1$ sessions $\{0,1,..,J\}$ share a GPS server of capacity R. Associated with the sessions is a set of parameters $f = (f_0, f_1,.., f_J)$, called the GPS assignment. The GPS assigns the instantaneous service rate to a backlogged session j as follows:

(12) $$r_j = R f_j \Big/ \sum_{i=0}^{J} f_i d_i$$

where $d_i = 1$ if session i is backlogged and $d_i = 0$ otherwise. Assuming GPS assignment: $f_0 = s^2$, $f_j = P_j x_j$ for $j = 1,..,J$ and that session 0 always generates traffic, DS-CDMA model (11) is identical to GPS (12). This mapping allows one to directly apply known performance results for GPS [4]-[6] to evaluate queueing performance of DS-CDMA at least in cases when approximation (11) is possible. In particular, model (11) with $R = R^{\max} = W/\ln 2$ gives upper limit on ability of DS-CDMA to provide Quality of Service (QoS) to multimedia traffic.

4. Examples

4.1. Approximation of Independent Queues for a Rigid DS-CDMA

In a very rigid DS-CDMA without retransmissions ($\Omega_j = 0$) the service bit rate for user j can be approximated by the following step-wise function:

(13) $\qquad c_j(SIR) = 0 \quad if \quad SIR \le SIR_j^*, \quad and \quad c_j(SIR) = \hat{c}_j \quad if \quad SIR > SIR_j^*$

It is easy to see that the approximation of independent queues can be used for performance evaluation of the model (13) under (7). User j gets fixed service rate $\hat{c}_j$ if the total interference at the base station $s^2 + P^+ < P_j x_j / SIR_j^*$ and gets no service otherwise. Without loss of generality assume that $P_1 x_1 / SIR_1^* \ge .. \ge P_J x_J / SIR_J^*$. It can be shown that the corresponding fixed point equation (8) may have multiple locally stable solutions describing system modes with users $j = 1,..,k$ receiving service at the peak rates $\hat{c}_j$ and completely blocked users $j = k+1,..,J$. In particular, the system persistent mode exists with all users receiving service at their peak rates is at least locally if $s^2 + R^c \sum P_j x_j \; l_j / N_j < P_J x_J / SIR_J^*$.

4.2. Approximation of Interacting Queues for a Flexible DS-CDMA

In this subsection we very briefly demonstrate how performance results for model (11) of an extremely flexible DS-CDMA can be derived from known results for GPS under various traffic conditions. First note that in a case of ideal power control user j is guaranteed the service rate g_j:

(14) $$c_j \ge g_j = \frac{P_j x_j}{s^2 + \sum P_i x_i} R,$$

regardless of transmission rates by other users. This allows the system even without scheduling and reservation to provide certain QoS guarantees and certain degree of fairness.

Given $l > 0$ and $b > 0$, the traffic stream is called $(l, b) - regular$, if the total amount of traffic $A(t_1, t_2)$ generated by the stream during any time interval $[t_1, t_2)$ satisfies the following condition: $A(t_1, t_2) \le b + l(t_2 - t_1)$. Parameters l and b upper bound the average rate and burstiness of the stream. Importance of $(l, b) - regular$ traffic streams is the result of the fact that these streams are the worst case scenario output of the leaky bucket regulator with token arrival rate l and the token buffer size b. For $(l, b) - regular$ traffic sources it is possible to provide worst-case deterministic bounds (i.e., hard guarantees) for delays and backlogs [4]-[5]. Session j generating $(l_j, b_j) - regular$ traffic stream is called locally stable if the average rate l_j does not exceed the session guaranteed service rate: $l_j \le g_j$. It is easy to see that for this locally stable session the maximum backlog and delay do not exceed b_j and $b_j / (g_j - l_j)$ respectively. Note that given g_j, these hard guarantees do not depend on sessions $i \ne j$. For locally unstable sessions where $l_j > g_j$ the hard guarantees for backlogs and delays can be derived if all the sessions $i = 1,..,J$ are $(l_i, b_i) - regular$ and the total average arrival rate does not exceed the capacity of the system: $\sum l_i \le R$. However in this case the hard guarantees depend on all session parameters. Since sources usually can tolerate certain losses, statistical (i.e., soft) guarantees for backlogs and delays are of great interest. Such soft guarantees have been derived in [ ] under the following traffic conditions: $\Pr\{A(t_1, t_2) \ge l(t_2 - t_1) + x\} \le \Lambda_j e^{-g_j x}$ for any t_1, t_2, x.

5. References

1. Gilhousen, K.,I. Jacobs, R. Padovani, A. Viterbi, L. Weaver, and C. Wheatley, III, "On the Capacity of a Cellula CDMA System," IEEE Trans. On Vehicular Technology, Vol. 40, No. 2, May 1991, pp. 303-312.
2. Marbukh V., and Moayeri N., "A Framework for Throughput and Stability Analysis of a DS-CDMA Network," IEEE Trans. On Vehicular Tech., May 1999.

3. Marbukh V., "Loss Circuit Switched Communication Networks: Performance Analysis and Dynamic Routing," Queueing Systems, 13(1993), 111-141.
4. Parekh, A.K. and Gallager, R.G., "A generalized Processor Sharing Approach to Flow Control in Integrated Services Network-The Single Node Case," IEEE/ACM Transactions on Networking, 1(3):344-357, 1993.
5. Parekh, A.K. and Gallager, R.G., "A generalized Processor Sharing Approach to Flow Control in Integrated Services Network-The Multiple Node Case," IEEE/ACM Transactions on Networking, 2(2):137-150, 1994.
6. Zhi-Li Zhang, Don Towsley, and Jim Kurose "Statistical Analysis of the Generalized Processor Sharing Scheduling Discupline," IEEE J. Select. Areas Commun., vol. 13, No 6, August 1995, , pp 1071-1080.